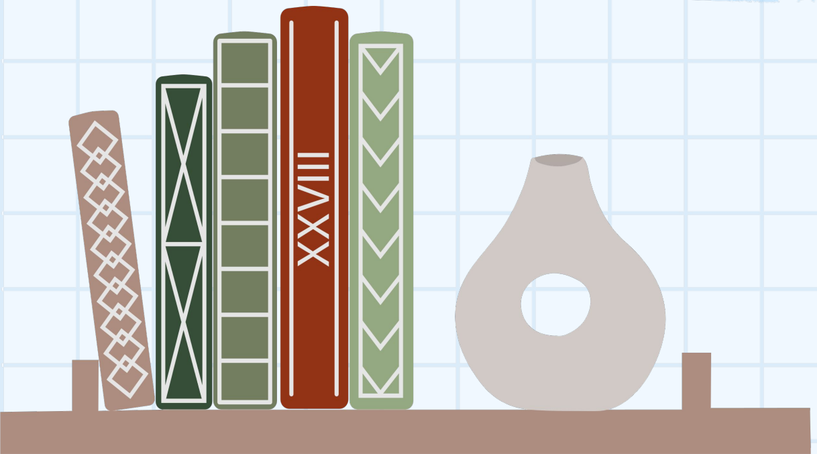
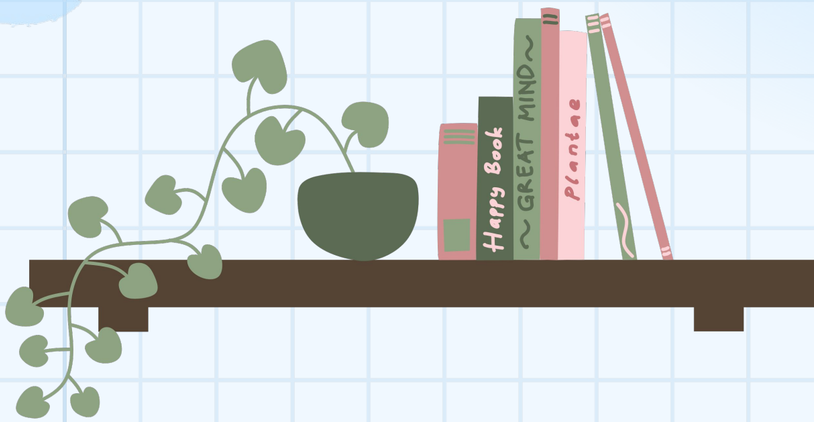
Powered by
Futurense

BS./BSC.IN

Applied AI and Data Science

Basics of Data Analytics





Powered by Futurense

Let's dive into and learn:

1 Removing Duplicates

Removing Duplicates

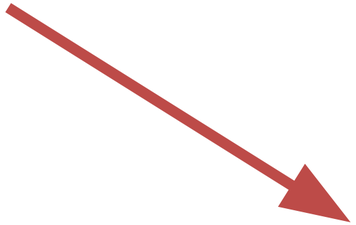
Powered by

Futurense

- Removing duplicates is usually the first task in the data cleaning process
- Could be due to human error
- Error during data entry
- While copying, saving or moving records
- Machine malfunctioning

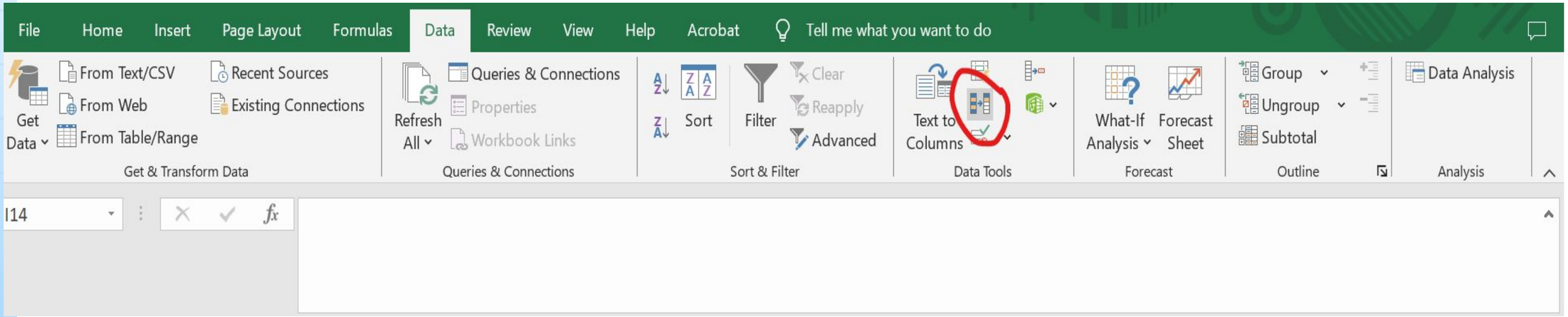
Removing duplicates

Transactic	Customer	Product ID	Purchase Date
53666	34987	453	01-02-2025
54333	35671	984	12-01-2025
56473	39823	382	29-01-2025
56234	36005	226	05-01-2025
54333	35671	984	12-01-2025



Transactic	Customer	Product ID	Purchase Date
53666	34987	453	01-02-2025
54333	35671	984	12-01-2025
56473	39823	382	29-01-2025
56234	36005	226	05-01-2025

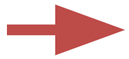
Removing Duplicates in Excel



Removing Duplicates

- Data may appear to be in duplicate because we are missing

Name	Age	Salary	Location
Priya	29	55000	Patna
Rahul	32	67000	Mohali
Devesh	28	52000	aneshwar
Bhavna	31	64000	Thrissur
Priya	29	55000	Patna



Name	Age	Salary	Location	DOB
Priya	29	55000	Patna	05-03-1996
Rahul	32	67000	Mohali	10-10-1993
Devesh	28	52000	aneshwar	12-07-1997
Bhavna	31	64000	Thrissur	13-08-1994
Priya	29	55000	Patna	01-04-1996

Removing Duplicates

Powered by

Futurense

- Before removing, ensure that you understand if the repeated data is indeed an error that we have to fix
- Make sure they are indeed duplicates

Recap

Powered by

Futurense

Removing duplicates

Powered by Futurense

Thank you